

(19)日本国特許庁（J P）

(12) 公 開 特 許 公 報（A）

(11)特許出願公開番号
特開平8－95596

(43)公開日 平成8年(1996)4月12日

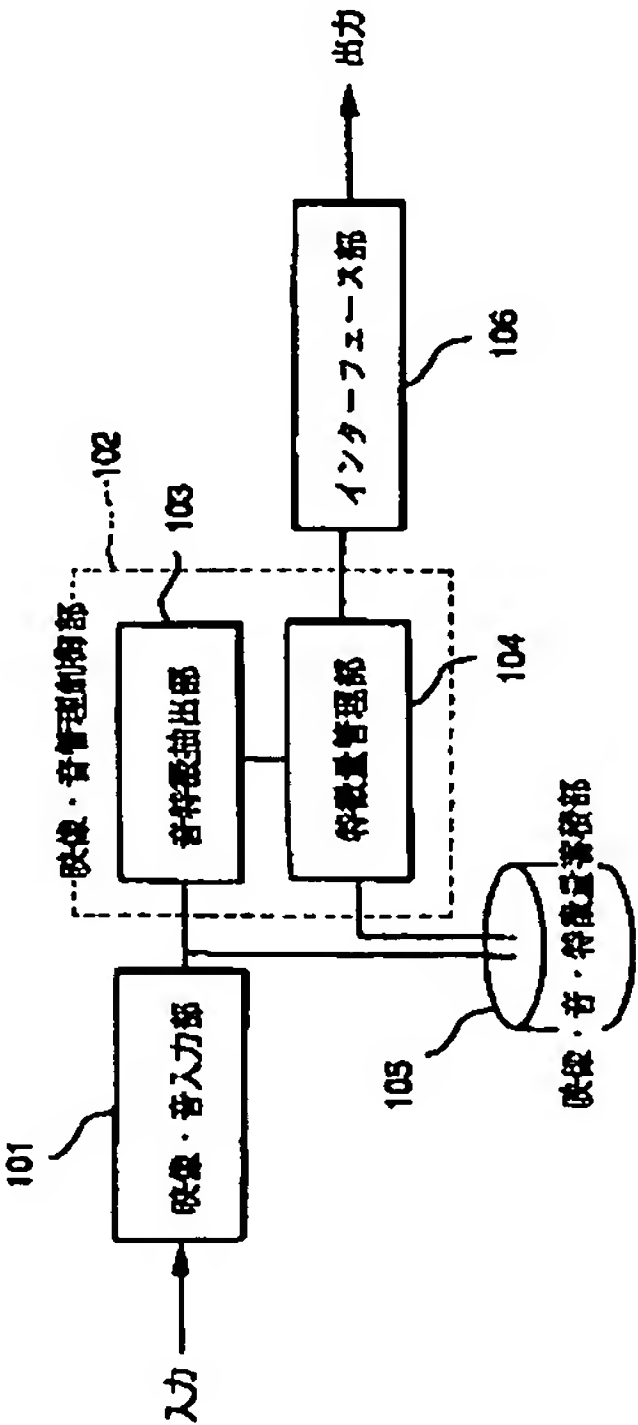
(51)IntCl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/02	A			
G 0 6 T 13/00				
G 1 1 B 20/00	Z	9294－5D 9365－5H	G 0 6 F 15/ 62 H 0 4 N 5/ 92	3 4 0 A E
審査請求 未請求 請求項の数23 O L （全 9 頁） 最終頁に続く				
(21)出願番号	特願平6－227545		(71)出願人	000004226 日本電信電話株式会社 東京都新宿区西新宿三丁目19番2号
(22)出願日	平成6年(1994)9月22日		(72)発明者	南 憲一 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
			(72)発明者	阿久津 明人 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
			(72)発明者	外村 佳伸 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
			(74)代理人	弁理士 志賀 富士弥

(54)【発明の名称】 速見速聴き装置及び方法

(57)【要約】

【目的】 音情報を用いて映像や音の外殻や雰囲気や短時間のうちに把握できる速見速聴き装置及び方法を提供する。

【構成】 まず、映像・入力部101から実時間で映像情報（画像情報と音情報から成る情報）或は音情報を入力し、実時間で入力された映像或いは音情報を映像・音・特徴量蓄積部105に蓄積し、蓄積された映像或いは音情報を入力する。次に、音特徴抽出部103により、入力された映像或いは音情報において、音情報から種々の特徴量を抽出する。次に、特徴量管理部104により、抽出された特徴量に基づいて映像或いは音情報を加工する。加工した映像或いは音情報はインターフェース部106より出力する。この特徴量に基づく加工によって、速見、速聴を実現する。



【特許請求の範囲】

【請求項 1】 実時間で画像情報と音情報からなる映像情報或は音情報を入力する映像・音情報入力部と、前記実時間で入力された映像或いは音情報を蓄積し該蓄積された映像或いは音情報を出力し該映像或いは音情報の特徴量を蓄積する映像・音・特徴量蓄積部と、前記入力された映像或いは音情報において音情報から種々の特徴量を抽出し該特徴量を管理する映像・音管理制御部と、前記特徴量の管理された映像或いは音情報を出力するインターフェース部と、を有することを特徴とする速見速聴き装置。

【請求項 2】 前記映像・音管理制御部は、音情報から種々の特徴量を抽出する音特徴抽出部と、該特徴量を管理し該特徴量に基づいて音情報を加工、変換する特徴量管理部とを有することを特徴とする請求項 1 記載の速見速聴き装置。

【請求項 3】 前記音特徴抽出部は、音情報から種々の特徴量を抽出しある条件を満たす特徴量が存在する音区間を音情報から検出する音情報変化検出手段を具備することを特徴とする請求項 2 記載の速見速聴き装置。

【請求項 4】 前記特徴量管理部は、一定の特徴量を有する音区間と映像或いは音の中で発生した事象とを対応付ける事象対応付け手段を具備することを特徴とする請求項 3 記載の速見速聴き装置。

【請求項 5】 前記特徴量管理部は、特定の特徴量を有する或いは特定の事象に対応する該音区間をつなぎ合わせるつなぎ合わせ手段を具備することを特徴とする請求項 3 または請求項 4 記載の速見速聴き装置。

【請求項 6】 前記特徴量管理部は、入力映像情報の場合には該音区間に対応する画像情報を同期させる再生手段を具備することを特徴とする請求項 3 または請求項 4 または請求項 5 記載の速見速聴き装置。

【請求項 7】 前記音情報変化検出手段は、音情報から周波数スペクトルを算出し、スペクトルの包絡線からピークを検出し、該ピークの特徴量がある条件を満たす音区間を音情報から検出するものであることを特徴とする請求項 3 記載の速見速聴き装置。

【請求項 8】 前記事象対応付け手段は、該ピークの特徴量を典型的な事象における該ピークの特徴量と比較することによって該音区間と事象を対応付けるものであることを特徴とする請求項 4 記載の速見速聴き装置。

【請求項 9】 前記つなぎ合わせ手段は、音情報に含まれる特定の特徴量或いは事象を有する音区間をつなぎ合わせ、再生した場合に要する時間の総和を該特徴量或いは該事象毎に算出し、再生時間の制限に応じて制限時間に最も近い音区間を選択する音区間選択手段を有することを特徴とする請求項 5 記載の速見速聴き装置。

【請求項 10】 前記つなぎ合わせ手段は、選択された該音区間において、音情報のパワーが低くなったところをつなぎ合わせ点とするものであることを特徴とする請

求項 5 または請求項 9 記載の速見速聴き装置。

【請求項 11】 前記つなぎ合わせ手段は、選択された該音区間において、該音区間の両端のパワーを低くしてつなぎ合わせ点とするものであることを特徴とする請求項 5 または請求項 9 記載の速見速聴き装置。

【請求項 12】 音情報変化抽出手段によって一定の特徴量或いはつなぎ合わせ手段によって対応づけられた特定の事象毎に分類された音区間の集まりを各々の振幅を周期的に変化させて再生する再生手段を有することを特徴とする請求項 4 または請求項 5 または請求項 6 記載の速見速聴き装置。

【請求項 13】 まず、実時間で画像情報と音情報からなる映像情報或いは音情報を入力し、前記実時間で入力された映像或いは音情報を蓄積し、実時間でない場合には蓄積された映像或いは音情報を入力し、次に、前記入力された映像或いは音情報において音情報から種々の特徴量を抽出し、次に、前記抽出した特徴量に基づいて前記映像或いは音情報を加工し、次に、前記加工した映像或いは音情報を出力することを特徴とする速見速聴き方法。

【請求項 14】 映像或は音情報を加工する過程において、音情報から抽出された種々の特徴量のうち、ある条件を満たす特徴量が存在する音区間を音情報から検出する音情報変化検出方法を用いることを特徴とする請求項 13 記載の速見速聴き方法。

【請求項 15】 一定の特徴量を有する音区間と映像或いは音の中で発生した事象とを対応付ける事象対応付け方法を用いることを特徴とする請求項 14 記載の速見速聴き方法。

【請求項 16】 特定の特徴量を有する或いは特定の事象に対応する該音区間をつなぎ合わせるつなぎ合わせ方法を用いることを特徴とする請求項 14 または請求項 15 記載の速見速聴き方法。

【請求項 17】 入力映像情報の場合には該音区間に対応する画像情報を同期させる再生方法を用いることを特徴とする請求項 14 または請求項 15 または請求項 16 記載の速見速聴き方法。

【請求項 18】 前記音情報変化検出方法であって、音情報から周波数スペクトルを算出し、スペクトルの包絡線からピークを検出し、該ピークの特徴量がある条件を満たす音区間を音情報から検出する方法を用いることを特徴とする請求項 14 記載の速見速聴き方法。

【請求項 19】 前記事象対応付け方法であって、該ピークの特徴量を典型的な事象における該ピークの特徴量と比較することによって該音区間と事象を対応付ける方法を用いることを特徴とする請求項 15 記載の速見速聴き方法。

【請求項 20】 前記つなぎ合わせ方法であって、音情報に含まれる特定の特徴量或いは事象を有する音区間をつなぎ合わせ、再生した場合に要する時間の総和を該特

3

微量或いは該事象毎に算出し、再生時間の制限に応じて制限時間に最も近い音区間を選択する音区間選択方法を用いることを特徴とする請求項 1 6 記載の速見速聴き方法。

【請求項 2 1】 前記つなぎ合わせ方法であって、選択された該音区間において、音情報のパワーが低くなったところをつなぎ合わせ点とする方法を用いることを特徴とする請求項 1 6 または請求項 2 0 記載の速見速聴き方法。

【請求項 2 2】 前記つなぎ合わせ方法であって、選択された該音区間において、該音区間の両端のパワーを低くしてつなぎ合わせ点とする方法を用いることを特徴とする請求項 1 6 または請求項 2 0 記載の速見速聴き方法。

【請求項 2 3】 音情報変化抽出方法によって一定の特徴量或いはつなぎ合わせ方法によって対応づけられた特定の事象毎に分類された音区間の集まりを各々の振幅を周期的に変化させて再生する再生方法を用いることを特徴とする請求項請求項 1 5 または請求項 1 6 または請求項 1 7 記載の速見速聴き方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、映像や音の内容を短時間に把握する場合等に好適な速見速聴き装置及び方法に関するものである。

【0002】

【従来の技術】映像情報の速見技術には、ビデオブラウザ（大場：“動画像一覧表示技術：ビデオブラウザ”，1994 テレビジョン学会全国大会，pp. 219-220，1991）がある。これは、映像を早送りしたり、映像の切れ目であるカット点の前後のみを再生したりする等の操作によって、映像を見る時間を短縮するものである。また、映像の一覧性を向上させ、カット点直後の静止画を空間的に並べてディスプレイや紙の上に印刷し、一目で映像の外殻を把握できるようにする技術

（外村ら：“Paper Video：映像情報の紙への定着”，1994 電子情報通信学会春季全国大会，A-315，1994）がある。

【0003】音を利用したものでは、ワードスポッティング（Lynn Wilcox et. al, “Wordspotting for Voice Editing and Audio Indexing, Georgia Institute of Technology”，1992）と呼ばれる方法が存在し、これはマイクから音声を入力、認識してその言葉と同じ言葉が出てくる場所をテープ等の蓄積媒体から検索するものである。

【0004】

【発明が解決しようとする課題】映像や音の内容を把握したり、どの様な映像や音が蓄えられているのを短時間

4

で知るためには、再生したり早送りしたりしなければならない。しかし、この様な方法では、所望の部分にたどり着いたり、全体の雰囲気をつかんだりするまでにかなりの時間を要する。これらの解決方法としては、前述したようなビデオブラウザも考えられるが、単純な早送りでは音が再生できなかったり、映像の構造的な特徴量にのみ依存しているので、選ばれたブラウズする画像が映像の中で起きた事象を必ずしも反映しているとは言えない。また、つなぎ目での音がぶつ切れになって不自然であったり、さらに時間短縮を行いたい場合に数多くのカットや画像からどの様に重要なものだけを選択するかという問題が生じる。前述のワードスポッティング技術では、単語が予め分かっているわけではないのでブラウジングの機能は全く考えられていない。

【0005】本発明は上記のような問題点を解決するためになされたものであり、その目的は、音情報を用いて映像や音の外殻や雰囲気を短時間のうちに把握できる速見速聴き装置及び方法を提供することにある。

【0006】さらに詳しくは、映像は音の中で起きた事象を反映した映像或いは音の区間の選択を可能にし、幾つかの音の区間を再生する際には各区間のつなぎ目をぶつ切れの少ない自然なものにして映像や音の外殻や雰囲気をより短時間で把握できるようにする技術を提供する。また、音を利用して複数のカットや画像を幾つかにまとめたり、音の特徴量と事象を対応付けることによって、利用者が欲しいと思った事象に応じた速見或いは速聴きをすることを可能にする技術を提供する。さらに、欲しいと思った事象がはっきりと分からなくても異なる事象に対応する音の区間をある間隔で入れ替えながら周期的に再生する事が可能で、ブラウジングにも適した技術を提供する。

【0007】

【課題を解決するための手段】上記の目的を達成するため、請求項 1 の発明では、実時間で画像情報と音情報からなる映像情報或は音情報を入力する映像・音情報入力部と、前記実時間で入力された映像或いは音情報を蓄積し該蓄積された映像或いは音情報を出し該映像或いは音情報の特徴量を蓄積する映像・音・特徴量蓄積部と、前記入力された映像或いは音情報において音情報から種々の特徴量を抽出し該特徴量を管理する映像・音管理制御部と、前記特徴量の管理された映像或いは音情報を出し出力するインターフェース部と、を有する速見速聴き装置を手段とする。

【0008】請求項 2 の発明では、前記映像・音管理制御部が、音情報から種々の特徴量を抽出する音特徴抽出部と、該特徴量を管理し該特徴量に基づいて音情報を加工、変換する特徴量管理部とを有する速見速聴き装置を手段とする。

【0009】請求項 3 の発明では、前記音特徴抽出部が、音情報から種々の特徴量を抽出しある条件を満たす

5

特徴量が存在する音区間を音情報から検出する音情報変化検出手段を具備する速見速聴き装置を手段とする。

【0010】請求項4の発明では、前記特徴量管理部が、一定の特徴量を有する音区間と映像或いは音の中で発生した事象とを対応付ける事象対応付け手段を具備する速見速聴き装置を手段とする。

【0011】請求項5の発明では、前記特徴量管理部が、特定の特徴量を有する或いは特定の事象に対応する該音区間をつなぎ合わせるつなぎ合わせ手段を具備する速見速聴き装置を手段とする。

【0012】請求項6の発明では、前記特徴量管理部が、入力映像情報の場合には該音区間に対応する画像情報を同期させる再生手段を具備する速見速聴き装置を手段とする。

【0013】請求項7の発明では、前記音情報変化検出手段が、音情報から周波数スペクトルを算出し、スペクトルの包絡線からピークを検出し、該ピークの特徴量がある条件を満たす音区間を音情報から検出するものである速見速聴き装置を手段とする。

【0014】請求項8の発明では、前記事象対応付け手段が、該ピークの特徴量を典型的な事象における該ピークの特徴量と比較することによって該音区間と事象を対応付けるものである速見速聴き装置を手段とする。

【0015】請求項9の発明では、前記つなぎ合わせ手段が、音情報に含まれる特定の特徴量或いは事象を有する音区間をつなぎ合わせ、再生した場合に要する時間の総和を該特徴量或いは該事象毎に算出し、再生時間の制限に応じて制限時間に最も近い音区間を選択する音区間選択手段を有する速見速聴き装置を手段とする。

【0016】請求項10の発明では、前記つなぎ合わせ手段が、選択された該音区間において、音情報のパワーが低くなったところをつなぎ合わせ点とするものである速見速聴き装置を手段とする。

【0017】請求項11の発明では、前記つなぎ合わせ手段が、選択された該音区間において、該音区間の両端のパワーを低くしてつなぎ合わせ点とするものである速見速聴き装置を手段とする。

【0018】請求項12の発明では、前記再生手段が、音情報変化抽出手段によって一定の特徴量或いはつなぎ合わせ手段によって対応づけられた特定の事象毎に分類された音区間の集まりを各々の振幅を周期的に変化させて再生する手段を有する速見速聴き装置を手段とする。

【0019】請求項13の発明では、まず、実時間で画像情報と音情報からなる映像情報或いは音情報を入力し、前記実時間で入力された映像或いは音情報を蓄積し、実時間でない場合には蓄積された映像或いは音情報を入力し、次に、前記入力された映像或いは音情報において音情報から種々の特徴量を抽出し、次に、前記抽出した特徴量に基づいて前記映像或いは音情報を加工し、次に、前記加工した映像或いは音情報を出力する速見速

6

聴き方法を手段とする。

【0020】請求項14の発明では、映像或いは音情報を加工する過程において、音情報から抽出された種々の特徴量のうち、ある条件を満たす特徴量が存在する音区間を音情報から検出する音情報変化検出方法を用いる速見速聴き方法を手段とする。

【0021】請求項15の発明では、一定の特徴量を有する音区間と映像或いは音の中で発生した事象とを対応付ける事象対応付け方法を用いる速見速聴き方法を手段とする。

【0022】請求項16の発明では、特定の特徴量を有する或いは特定の事象に対応する該音区間をつなぎ合わせるつなぎ合わせ方法を用いる速見速聴き方法を手段とする。

【0023】請求項17の発明では、入力映像情報の場合には該音区間に対応する画像情報を同期させる再生方法を用いる速見速聴き方法を手段とする。

【0024】請求項18の発明では、前記音情報変化検出方法であって、音情報から周波数スペクトルを算出し、スペクトルの包絡線からピークを検出し、該ピークの特徴量がある条件を満たす音区間を音情報から検出する方法を用いる速見速聴き方法を手段とする。

【0025】請求項19の発明では、前記事象対応付け方法であって、該ピークの特徴量を典型的な事象における該ピークの特徴量と比較することによって該音区間と事象を対応付ける方法を用いる速見速聴き方法を手段とする。

【0026】請求項20の発明では、前記つなぎ合わせ方法であって、音情報に含まれる特定の特徴量或いは事象を有する音区間をつなぎ合わせ、再生した場合に要する時間の総和を該特徴量或いは該事象毎に算出し、再生時間の制限に応じて制限時間に最も近い音区間を選択する音区間選択方法を用いる速見速聴き方法を手段とする。

【0027】請求項21の発明では、前記つなぎ合わせ方法であって、選択された該音区間において、音情報のパワーが低くなったところをつなぎ合わせ点とする方法を用いる速見速聴き方法を手段とする。

【0028】請求項22の発明では、前記つなぎ合わせ方法であって、選択された該音区間において、該音区間の両端のパワーを低くしてつなぎ合わせ点とする方法を用いる速見速聴き方法を手段とする。

【0029】請求項23の発明では、音情報変化抽出方法によって一定の特徴量或いはつなぎ合わせ方法によって対応づけられた特定の事象毎に分類された音区間の集まりを各々の振幅を周期的に変化させて再生する再生方法を用いる速見速聴き方法を手段とする。

【0030】

【作用】本発明の請求項1及び2及び13の速見速聴き装置及び方法は、実時間で映像情報（画像情報と音情報

から成る情報)、音情報を入力し、実時間で入力された映像或いは音情報を蓄積し、蓄積された映像或いは音情報を入力し、入力された映像或いは音情報において、音情報から種々の特徴量を抽出し、その特徴量に基づいて映像或いは音情報を管理/加工し、管理/加工した映像或いは音情報を出力する手段を用いることにより、実時間情報、蓄積情報によらず映像或いは音情報を入力し、管理/加工し、出力し、その特徴量に基づく管理/加工によって速見、速聴きを実現する。

【0031】本発明の請求項3の音特徴抽出部は、音情報から種々の特徴量を抽出し、ある条件を満たす特徴量が存在する音区間を音情報から検出する手段を用いることにより、音情報を分割できるようにする。

【0032】本発明の請求項4の特徴量管理部は、一定の特徴量を有する音区間と映像或いは音の中で発生した事象とを対応付ける手段を用いることにより、音情報を分類できるようにする。

【0033】本発明の請求項5の特徴量管理部は、特定の特徴量を有する或いは特定の事象に対応する該音区間をつなぎ合わせる手段を用いることにより、所望の音区間をまとめる。

【0034】本発明の請求項6の特徴量管理部は、入力映像情報の場合には該音区間に対応する画像情報を同期させる手段を用いることにより、画像と音がずれることなく再生する。

【0035】本発明の請求項14の速見速聴き方法は、音情報から種々の特徴量を抽出し、ある条件を満たす特徴量が存在する音区間を音情報から検出する方法を用いることにより、音情報を分割できるようにする。

【0036】本発明の請求項15の速見速聴き方法は、一定の特徴量を有する音区間と映像或いは音の中で発生した事象とを対応付ける方法を用いることにより、音情報を分類できるようにする。

【0037】本発明の請求項16の速見速聴き方法は、特定の特徴量を有する或いは特定の事象に対応する該音区間をつなぎ合わせる方法を用いることにより、所望の音区間をまとめる。

【0038】本発明の請求項17の速見速聴き方法は、入力映像情報の場合には該音区間に対応する画像情報を同期させる方法を用いることにより、画像と音がずれることなく再生する。

【0039】本発明の請求項7、18の音情報変化検出手段及び方法は、音情報から周波数スペクトルを算出し、スペクトルの包絡線からピークを検出し、該ピークの特徴量がある条件を満たす音の区間を音情報から検出する方法を用いることにより、音の変化を検出して特徴を検出する。

【0040】本発明の請求項8、19の事象対応付け手段及び方法は、該ピークの特徴量を典型的な事象における該ピークの特徴量と比較することによって該音区間と

事象を対応付ける。

【0041】本発明の請求項9、20のつなぎ合わせ手段及び方法は、音情報に含まれる特定の特徴量或いは事象を有する音区間をつなぎ合わせ、再生した場合に要する時間の総和を該特徴量或いは該事象毎に算出し、再生時間の制限に応じて制限時間に最も近い音区間を選択する手段を用いることにより、再生時間を変化指せ、制限時間に応じた速見速聴きを実現する。

【0042】本発明の請求項10、21のつなぎ合わせ手段及び方法は、音情報のパワーが低くなったところをつなぎ合わせ点とする方法を用いることにより、選択された該音区間をつなぎ合わせる際に生じる音の不自然なぶつ切れをなくす。

【0043】本発明の請求項11、22のつなぎ合わせ手段及び方法は、該音区間の両端のパワーを低くしてつなぎ合わせ点とする方法を用いることにより、選択された該音区間のつなぎ合わせる際に生じる音の不自然なぶつ切れをなくす。

【0044】本発明の請求項12、23の再生手段及び方法は、前記音情報変化抽出手段によって一定の特徴量或いは前記つなぎ合わせ手段によって対応づけられた特定の事象毎に分類された音区間の集まりを各々の振幅を周期的に変化させる手段を用いることにより、異なる特徴や事象に対応した音をある間隔で順番に再生し、どのような特徴或は事象が含まれているのかを短時間で把握可能とする。

【0045】

【実施例】次に、本発明の一実施例について図面を参照して説明する。

【0046】図1は本発明の一実施例の速見速聴き装置の概略構成を示すブロック図である。本実施例の速見速聴き装置は、映像或いは音情報を入力する映像・音入力部101と、実時間で入力された映像、音、特徴量を逐次蓄積する映像・音・特徴量蓄積部105と、映像、音、特徴量を管理し、制御する映像・音管理制御部102と、出力された映像、音を提示するインターフェース部106から構成されている。映像・音管理制御部102は、音情報の特徴を抽出する音特徴抽出部103と、特徴量に基づいて映像或いは音を加工、変換、再生する特徴量管理部104で構成されており、音特徴抽出部103と特徴量管理部104は各々、並列或いは時分割で作動し、実時間で特徴抽出しながら映像或いは音を加工、変換、再生できる。

【0047】図2は映像・音管理制御部102の構成を示すブロック図である。音特徴抽出部103は、音の特徴がある条件を満たす音の区間を検出する音情報変化検出手段201から成り、特徴量管理部104は、音情報変化検出手段201によって求められた音区間と映像或いは音の中で起きた事象を対応付ける事象対応付け手段202と、該音区間をつなぎ合わせるつなぎ合わせ手段

203と、該音区間と対応する画像情報を同期させる再生手段204によって構成されている。

【0048】図3は、映像・音管理制御部102の音特徴抽出部103及び特徴量管理部104を計算機等のソフトウェア的に実現する場合の処理を示すフローチャートである。この場合、まず、音情報変化検出処理301を行い、次に、事象対応付け処理302を行い、次に、つなぎ合わせ処理303を行い、次に、再生処理304を行う。

【0049】次に、本実施例の動作を説明する。

【0050】図4は音情報変化検出手段201または音情報変化検出処理301の処理を示すフローチャート、図5は本実施例による音の波形から周波数スペクトルとスペクトルの包絡線を算出した様子を示す図、図6は本実施例によるスペクトル包絡のピークや特徴量を示す図、図7は本実施例によるつなぎ合わせ手段において用いられる波形のパワーの時間変化を示す図、図8は本実施例による再生手段において用いられる重み付け関数を示す図である。

【0051】映像或いは音情報は映像・音入力部101によって入力され、映像或いは音が実時間で入力された場合には、逐次映像・音・特徴量蓄積部105によって蓄積される。入力された映像情報は画像情報と音情報に分離され、音情報は音情報変化検出手段201によって解析される。入力が音情報のみの場合にはそのまま音情報変化検出手段201によって解析される。音情報変化検出手段201では、始めに図4のスペクトル算出処理401が施される。入力された音情報の波形が図5の501の様に与えられたとすれば、502の様に数ms～数10msのフレームに切り出され、FFT（高速フーリエ変換）処理が施される。FFT処理を施された波形は503に示すようにパワー方向の起伏が激しいのでピーク検出には不適當である。そこで、図4の周波数スペクトル包絡線算出処理402を施して、FFTケプストラムを求め、リフタリング処理を施した後、もう一度FFT処理を施してスペクトルの包絡線を算出する。算出された包絡線は、504のようになる。ケプストラムの算出方法にはFFTの他にLPC（線形予測分析）を用いる方法も考えられる。

【0052】次に、求められた包絡線から図4の403に示すピーク検出処理を施す。リフタリング処理によって包絡線がなめらかになっているため包絡線の極大値及び極小値から容易にピークを求めることができる。求められたピークの例を図6の604～608に示す。

【0053】音情報の波形から切り出すフレームは数ms～数10ms程度時間軸方向にずらし、同様の処理を繰り返し行う。この様にして各時刻において求められたピークを基にピーク変動算出処理404が施される。この処理では、ピークを基準とした種々のパラメータの時間的な変化を求める。パラメータとしては、図6に示す

ピークのパワー601、ピークの周波数602、極大ピークから隣接する極小ピークまでのパワーの差603、ピークの数等が考えられる。

【0054】これらのパラメータの時間変化から事象対応付け手段202によって映像や音の中で起きた事象との対応付けが行われる。例えば、映像や音に音楽が含まれていたとすれば、ピークの周波数602の時間変動がある閾値以下であったり、一定周期でピークのパワー601が大きくなる等の特徴が見られるので、この様な特徴を含む音区間は音楽と対応付けられる。また、ピークの周波数がなめらかに変化したり、極大ピークと極小ピークの差が大きい等の特徴が見られる場合には人の声と対応付けられ、ピークのパワーや数が細かく変動する場合には雑踏や雑音と対応付けられる。

【0055】次につなぎ合わせ手段203により、例えば人の声の区間だけを再生したい場合には事象対応付け手段202によって対応づけられた人の声の音区間が選択されつなぎ合わせられる。この際、図7の702が選択された音区間の1つであるとする、703、704の様に音区間の始めと終わりの近傍においてパワーが低くなる部分をつなぎ合わせるにより、不自然な音のぶつ切れを軽減することができる。また、音区間702の両端のパワーを強制的に低くしてつなげる方法もある。パワーの算出には音波形を数msの範囲で区切り振幅の2乗和を用いる方法等が考えられる。算出されたパワーの一例を701に示す。

【0056】再生する音区間は一定の特徴量をもった区間のつなぎ合わせだけでなく、音情報が急激に変化した点の前後一定時間にすることも可能である。

【0057】つなぎ合わせ手段203によって生成された音区間の集まりは再生手段204によって再生される。この際、入力情報が映像であった場合には、各々の音区間に対応した画像と共に再生される。入力情報が音のみの場合にはそのまま再生される。

【0058】再生手段には次の様なブラウジング機能も付いている。一定の特徴量或いは特定の事象毎につなぎ合わされた音区間に、各々振幅方向に図8の801～804の様な周期的に変化する重み付けを施し、805に示すような位相差を与えて再生する。図8は音区間の分類が4つの場合を示しているが、音区間の数や重み付け関数、位相差等は任意に設定できる。この様な方法で再生することによって異なる特徴或いは事象に対応した音区間が入れ替わりながら聞こえてくる。

【0059】

【発明の効果】以上説明したように、本発明は以下に示すような効果がある。

【0060】請求項1～12の発明は、音情報からその特徴をとらえ、映像や音の中で起きた事象との対応付けを行い、所望の事象や特徴量がある条件を満たす音の区間を再生するため、映像或いは音の内容を反映した速見

10

20

30

40

50

或いは速聴きができる。

【0061】請求項9、20のつなぎ合わせ手段方法は、音情報に含まれる特定の特徴量或いは事象を有する音区間をつなぎ合わせ、再生した場合に要する時間の総和を特徴量或いは該事象毎に算出し、利用者が指定した再生制限時間に最も近い音区間を選択するため、再生時間の制限に応じた速見或いは速聴きができる。

【0062】本発明の請求項10、21のつなぎ合わせ手段及び方法は、選択された該音区間をつなぎ合わせる際に音情報のパワーが低くなったところをつなぎ合わせ点とするため、不自然な音のぶつ切れを軽減することができる。

【0063】本発明の請求項11、22のつなぎ合わせ手段及び方法は、選択された音区間をつなぎ合わせる際に、音区間の両端のパワーを低くしてつなぎ合わせるため、不自然の音のぶつ切れを軽減することができる。

【0064】本発明の請求項12、23の再生手段及び方法は、一定の特徴量或いは特定の事象毎に分類された音区間の集まりを各々の振幅を周期的に変化させて再生するため、どの様な特徴量或いは事象が含まれているのかを短時間で把握する事が可能になる。

【図面の簡単な説明】

【図1】本発明の一実施例の速見速聴き装置の概略構成を示すブロック図である。

【図2】本発明の一実施例の映像・音管理制御部の構成を示すブロック図である。

【図3】本発明の一実施例の音特徴抽出部及び特徴量管理部の動作を計算機等でソフトウェア的に実現した場合の処理の流れを示すフローチャートである。

【図4】本実施例による音情報の周波数スペクトル算出処理から周波数スペクトルの包絡線のピーク変動を算出するまでの処理の流れを説明するためのフローチャート

である。

【図5】本実施例による音情報の波形の例とFFTによって算出された周波数スペクトル及びケプストラムからリフタリング処理を施しFFTによって算出された周波数スペクトルの包絡線の様子を示す図である。

【図6】本実施例による音情報の周波数スペクトルの包絡線と音情報の変化をとらえるために用いる各種パラメータを示す図である。

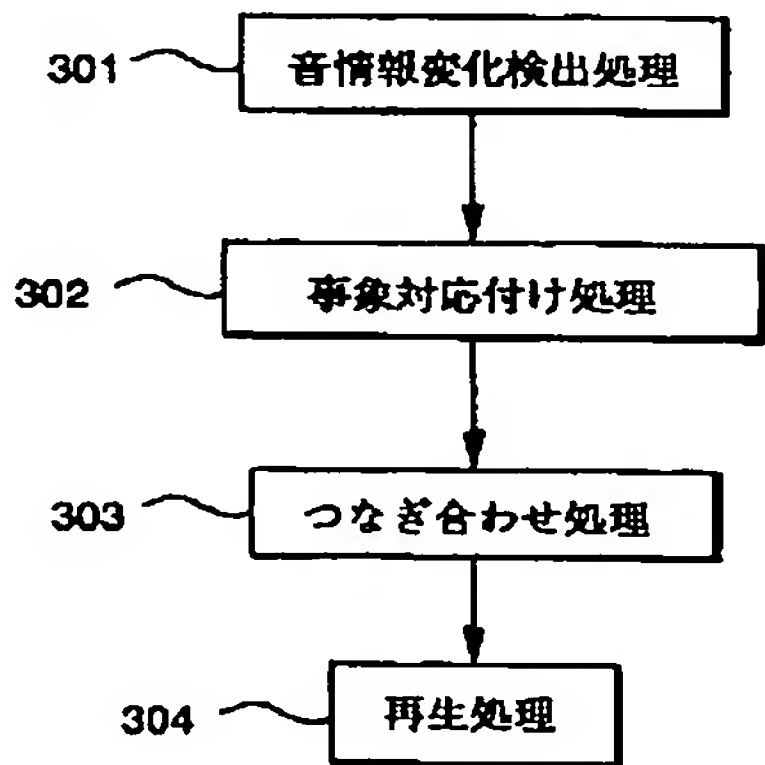
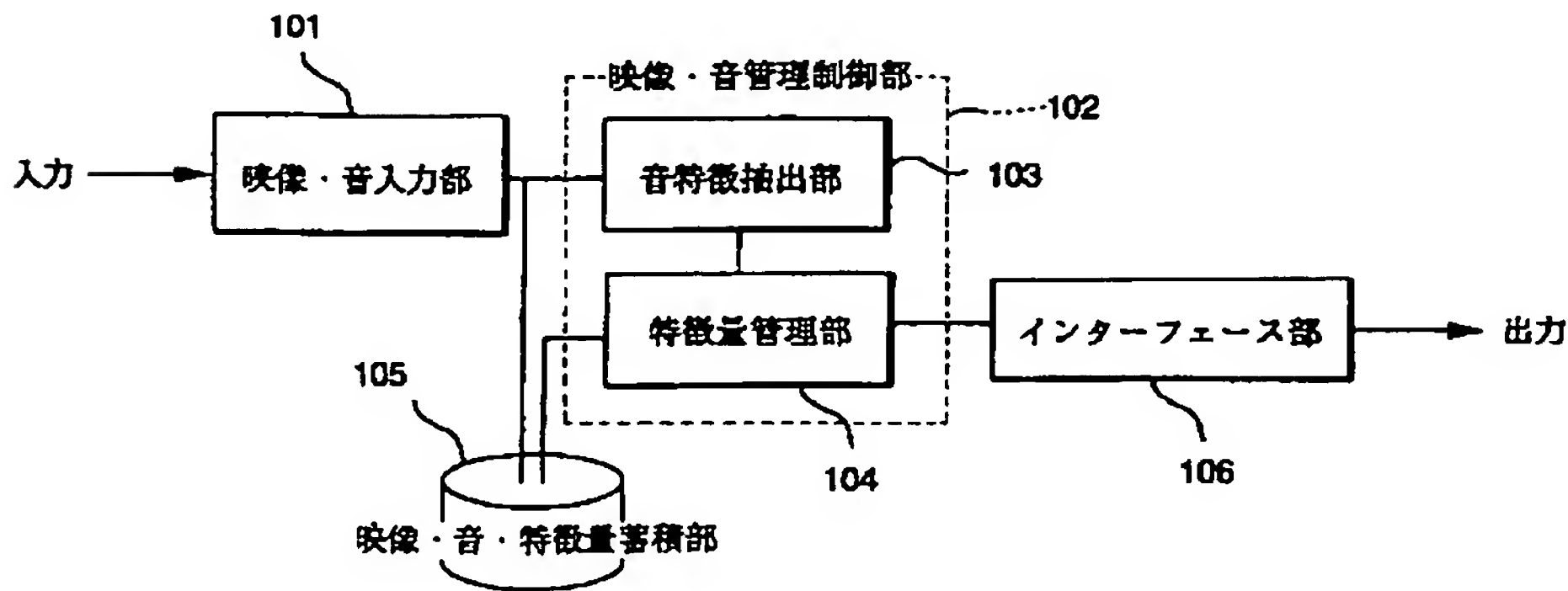
【図7】本実施例による音情報のパワーの様子とパワーが低くなったつなぎ合わせ部分を示す図である。

【図8】本実施例による音区間の重み付け関数と位相差の様子を示す図である。

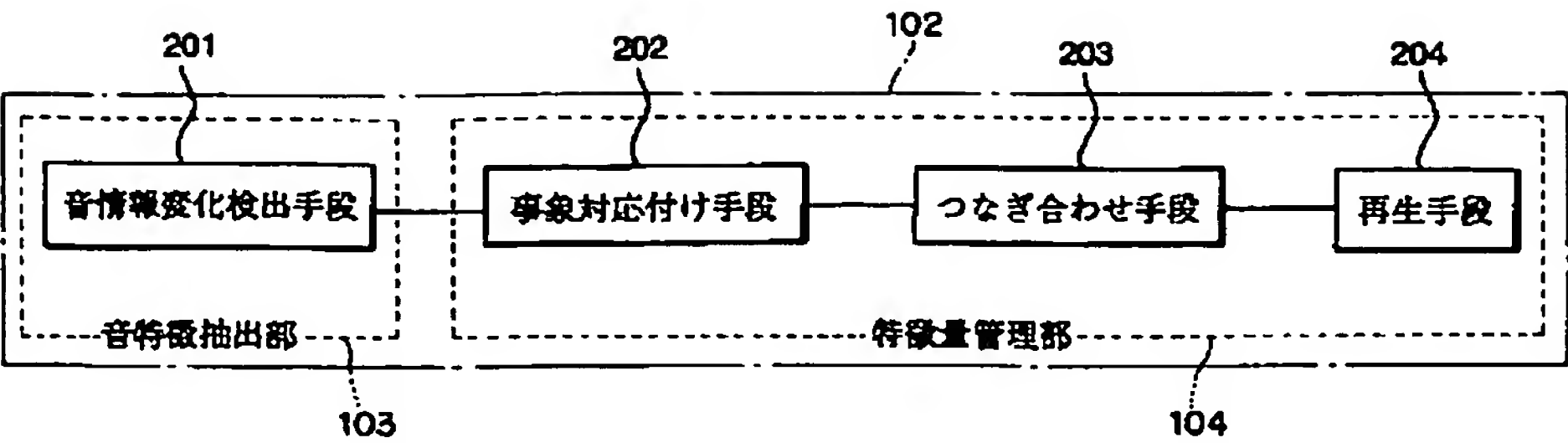
- 【符号の説明】
- 101…映像・音入力部
 - 102…映像・音管理制御部
 - 103…音特徴抽出部
 - 104…特徴量管理部
 - 105…映像・音・特徴量蓄積部
 - 106…インターフェース部
 - 201…音情報変化検出手段
 - 202…事象対応付け手段
 - 203…つなぎ合わせ手段
 - 204…再生手段
 - 301…音情報変化検出処理
 - 302…事象対応付け処理
 - 303…つなぎ合わせ処理
 - 304…再生処理
 - 401…周波数スペクトル算出処理
 - 402…周波数スペクトル包絡線算出処理
 - 403…ピーク検出処理
 - 404…ピーク変動算出処理

【図1】

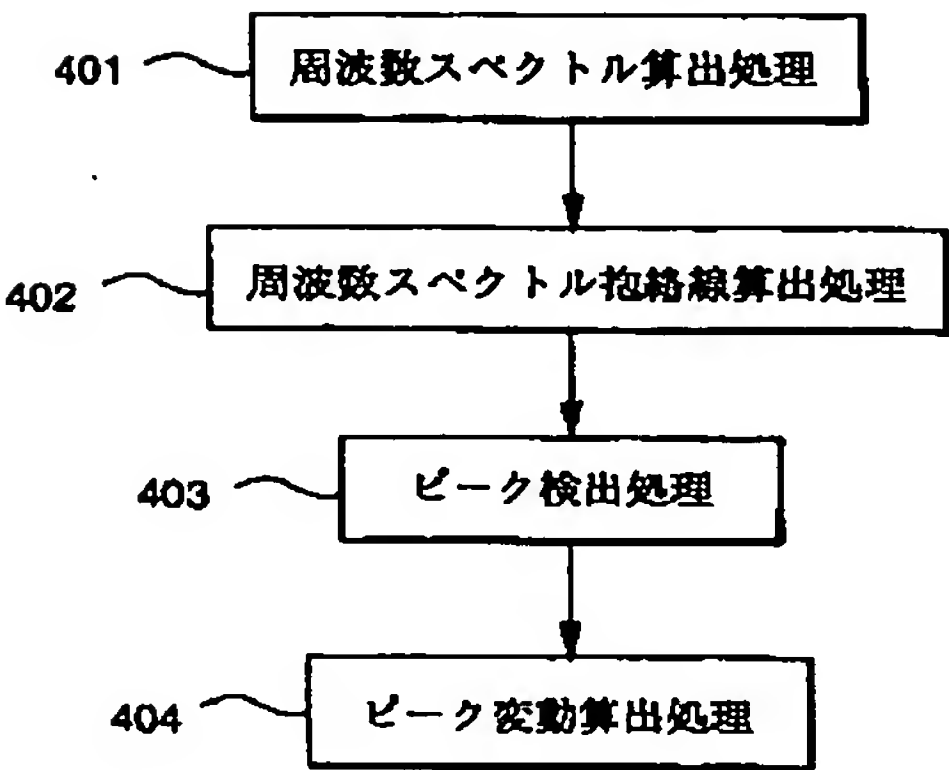
【図3】



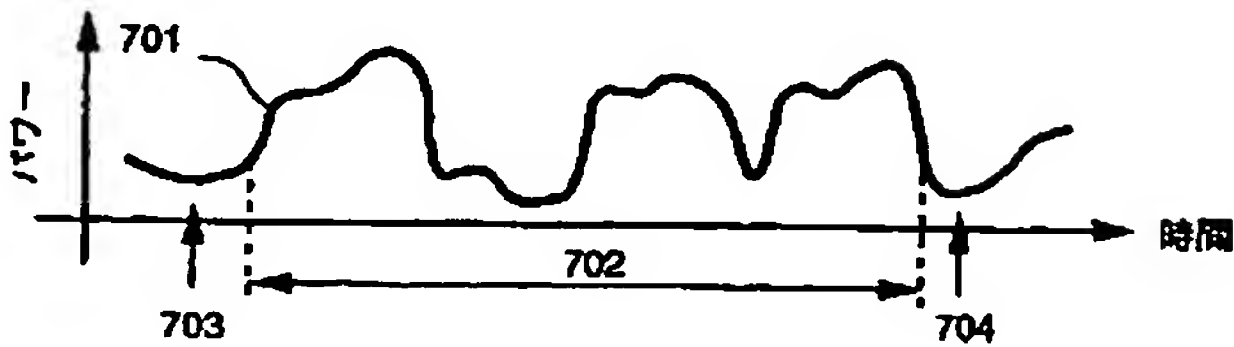
【図 2】



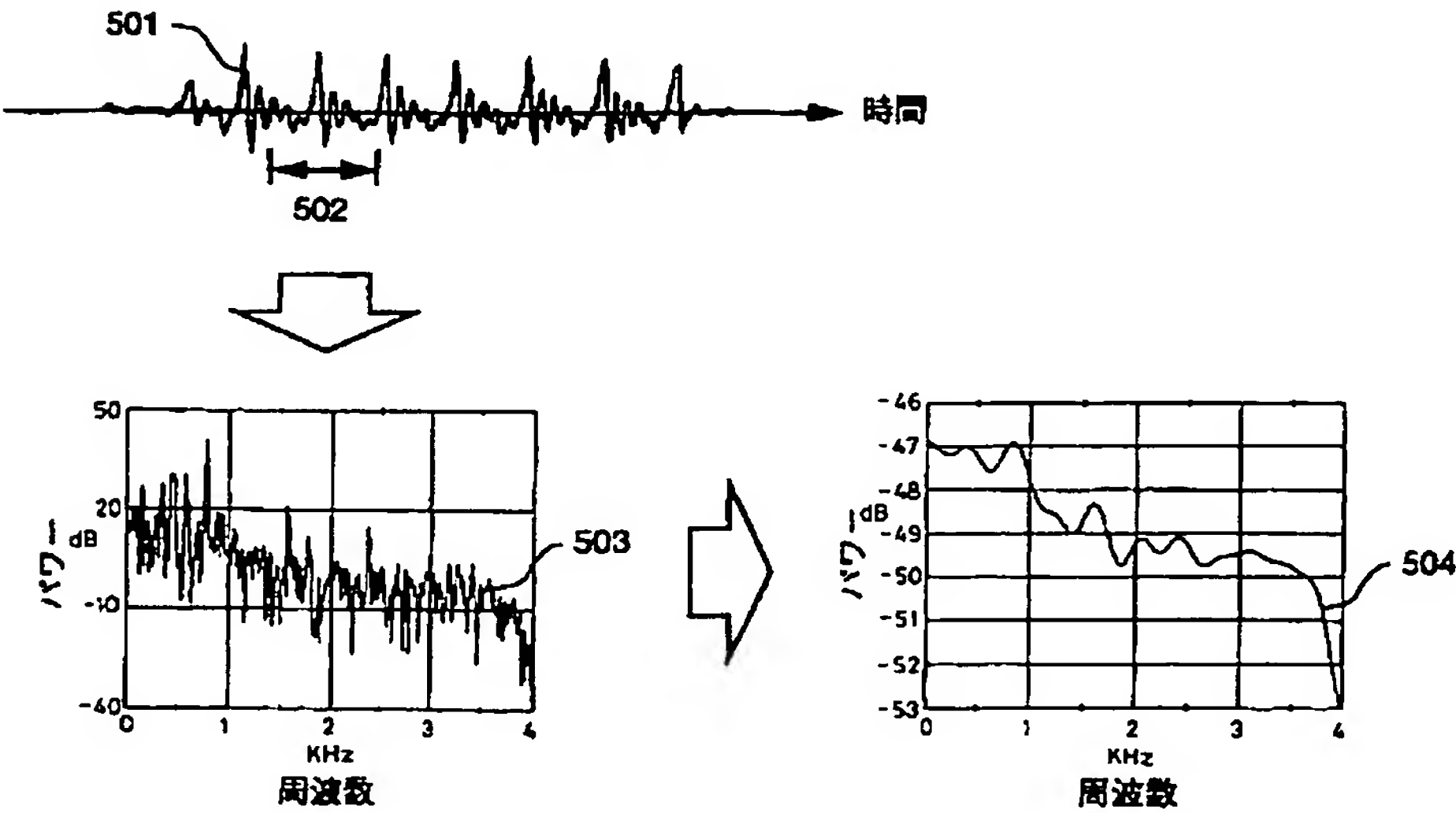
【図 4】



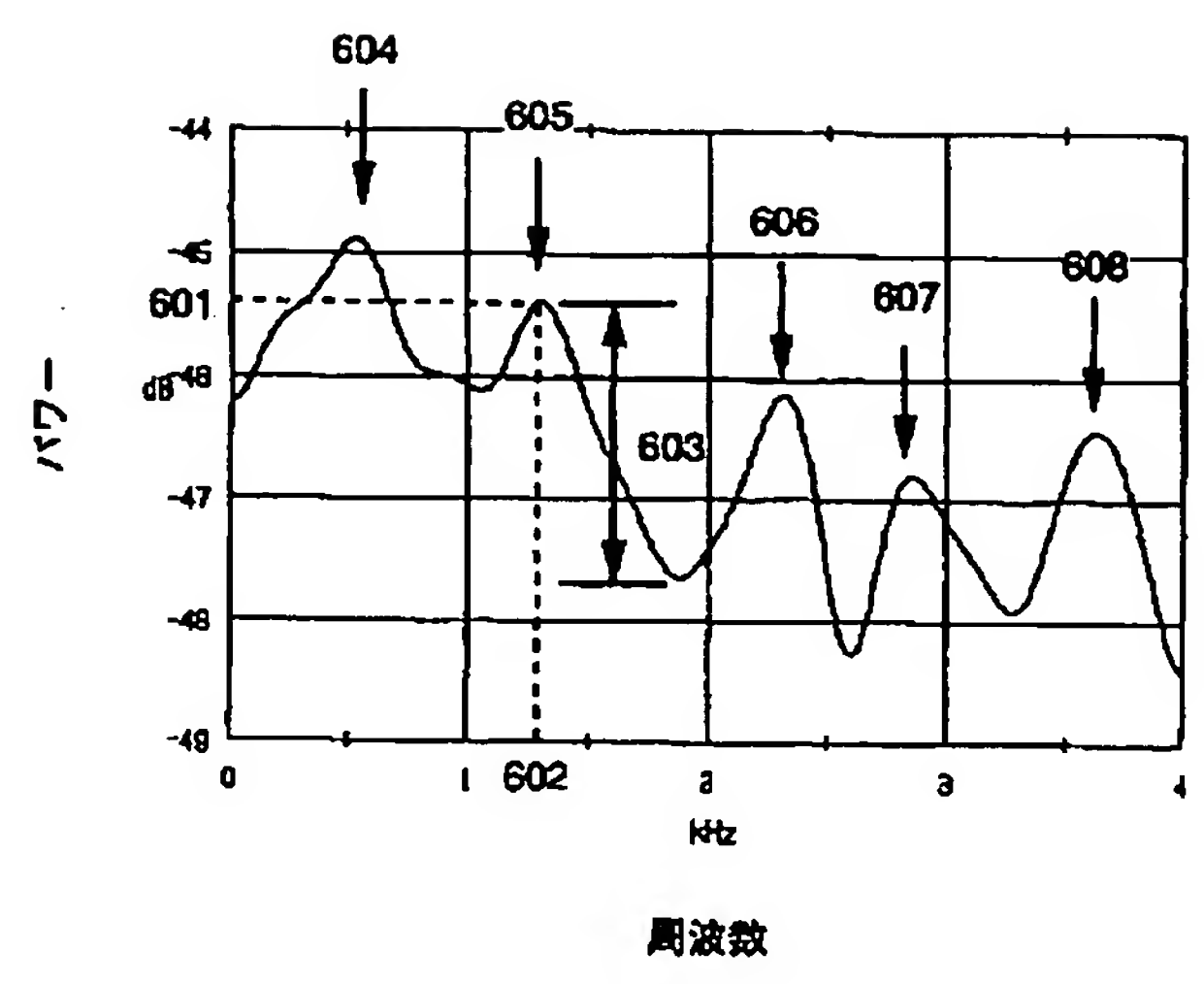
【図 7】



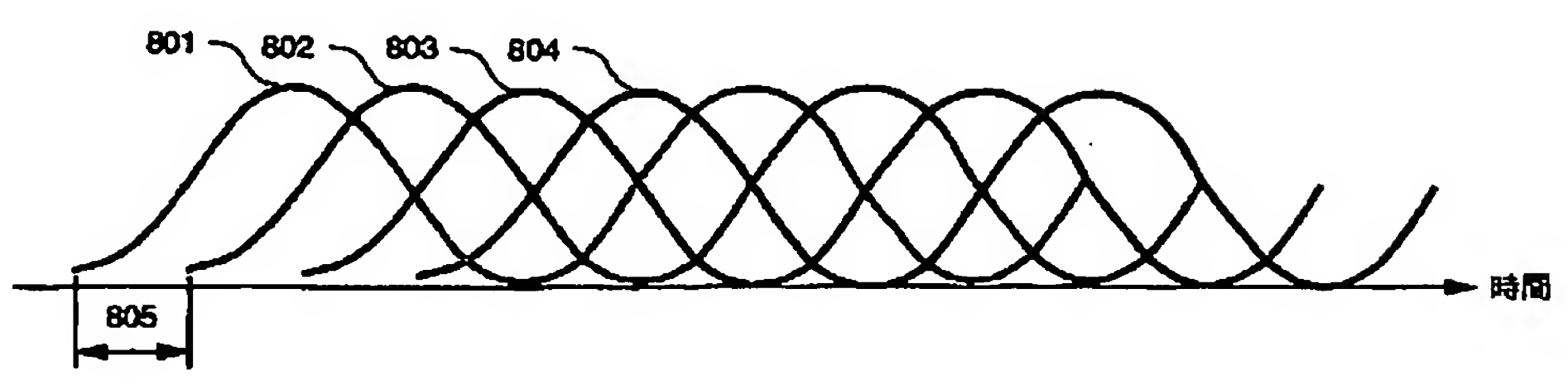
【図 5】



【図 6】



【図 8】



フロントページの続き